НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ

П. А. Меньшаков

Учреждение образования «Гомельский государственный технический университет имени П. О. Сухого», Беларусь

Научный руководитель И. А. Мурашко

Первоначальным этапом голосовой идентификации является получение голоса пользователя. Для этого необходим микрофон, фильтр и аналого-цифровой преобразователь для дальнейшей работы с цифровой записью голоса.

В общем виде процесс ввода речевых сообщений приведен на рис. 1.

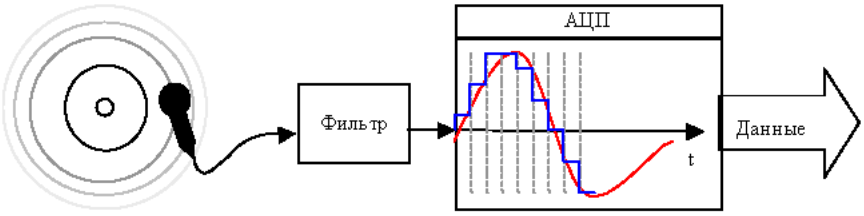

Рис. 1. Схема ввода записи голоса

С выхода микрофона сигнал подается на вход блока фильтрации. Следующим этапом является прохождение АЦП.

Далее оцифрованный сигнал попадает в блок цифровой обработки. В блоке цифровой обработки сигнал фильтруется и преобразуется в вектор, с которым в дальнейшем будет работать микропроцессор и нейросетевой обработчик.

Полученный вектор заносится в энергонезависимую память. Это необходимо для последующего сравнения с полученным отпечатком.

После сравнения отпечатка в памяти с полученным отпечатком микроконтроллер подает команду на блок управления внешним устройством, к примеру, на магнитный дверной замок. Общая схема устройства представлена на рис. 2.

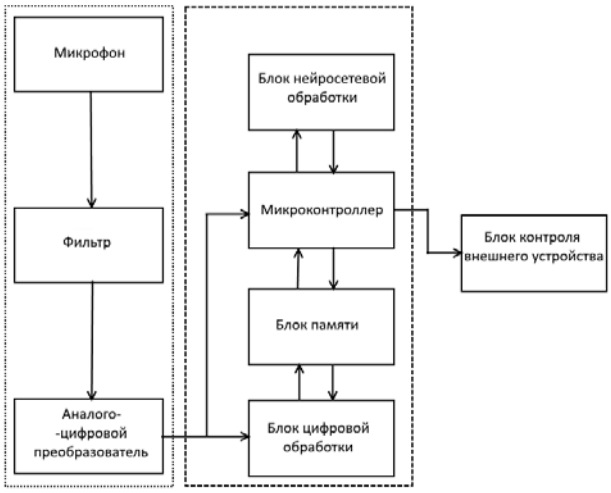

Рис. 2. Схема устройства

Сам процесс голосовой идентификации не требователен к ресурсам и состоит из двух этапов. Первый – получить голосовой отпечаток. Вторым шагом является сравнение голосовых отпечатков при помощи обученной нейронной сети. Для реализации процесса преобразования необходимо произвести определенный порядок действий.

При помощи микрофона получается запись голоса идентифицируемого и отправляется на ЭВМ. Наиболее оптимальным является получение WAV файла ввиду простоты работы с ним.

Полученную запись голоса необходимо разделить на кадры. Данное действие необходимо для более простой работы с записанной звуковой дорожкой.

Следующим этапом является устранение нежелательных эффектов и шумов. Это необходимо для того, чтобы записи, полученные в разное, время соответствовали друг другу независимо от сторонних факторов. Мною использовалось умножение каждого кадра на особую весомую функцию «Окно Хемминга»:

$$\omega(n) = 0{,}53836 - 0{,}46164 \cdot \cos\left(\frac{2\pi n}{N-1}\right), \tag{1}$$

где n – порядковый номер элемента в кадре, для которого вычисляется новое значение амплитуды; N – длина кадра (количество значений сигнала, измеренных за период).

Полученные кадры преобразуются в их частотную характеристику при помощи прогонки через «Быстрое Преобразование Фурье»:

$$X_k = \sum_{i=0}^{N-1} x_n e^{\frac{2\pi i}{N} kn}, \tag{2}$$

где N – длина кадра (количество значений сигнала, измеренных за период); xn – амплитуда n-го сигнала; Xk – N-комплексных амплитуд синусоидальных сигналов, слагающих исходный сигнал.

На сегодняшний день наиболее успешными являются системы распознавания голоса, использующие знания об устройстве слухового аппарата. Ввиду данных особенностей необходимо привести частотную характеристику каждого кадра к «мелам».

Для перехода к «мел»-характеристике используется следующая зависимость:

$$m = 1127 \log_e\left(1 + \frac{f}{700}\right), \tag{3}$$

где m – частота в мелах; f – частота в герцах.

Это последнее действие, необходимое для последующего преобразования в вектор характеристики, который впоследствии сравнивается с базой голосовых записей. Вектор будет состоять из мел-кепстральных коэффициентов, получить которые можно по следующей формуле:

$$C_n = \sum_{k=1}^{K} (\log S_k)\left[n\left(k - \frac{1}{2}\right)\frac{\pi}{K}\right], \tag{4}$$

где C_n – мел-кепстральный коэффициент под номером n; S_k – амплитуда k-го значения в кадре в мелах; K – наперед заданное количество мел-кепстральных коэффициентов $n \in [1, K]$.

Полученный вектор характеристик добавляется в базу данных для последующего сравнения с ним.

В работе использовалась нейронная сеть с обучением без учителя, так как оно является намного более правдоподобной моделью обучения в биологической системе. Процесс обучения выделяет статистические свойства обучающего множества и группирует сходные векторы в классы. Предъявление на вход вектора из данного класса даст определенный выходной вектор [3]. Схематически данная сеть изображена на рис. 3.

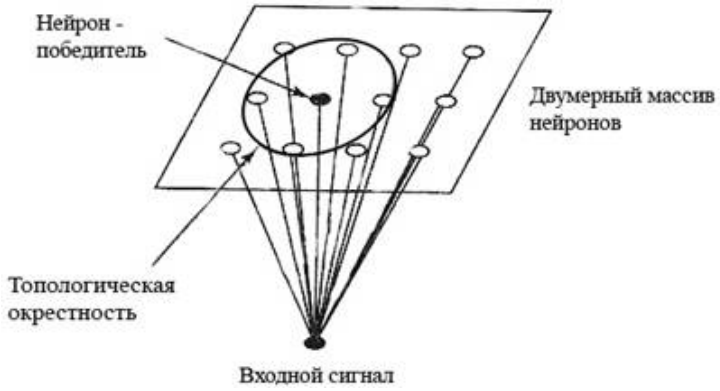

Рис. 3. Схема сети Кохонена

Распространение сигнала в такой сети происходит следующим образом: входной вектор нормируется на 1.0 и подается на вход, который распределяет его дальше через матрицу весов *W*. Каждый нейрон в слое Кохонена вычисляет сумму на своем входе и в зависимости от состояния окружающих нейронов этого слоя становится активным или неактивным (1.0 и 0.0). Нейроны этого слоя функционируют по принципу конкуренции, т. е. в результате определенного количества итераций активным остается один нейрон или небольшая группа. Так как отработка этого механизма требует значительных вычислительных ресурсов, в моей модели он заменен нахождением нейрона с максимальной активностью и присвоением ему активности 1.0, а всем остальным нейронам 0.0.

Если сеть находится в режиме обучения, то для выигравшего нейрона происходит коррекция весов матрицы связи по формуле

$$w_{\text{н}} = w_{\text{м}} + a(x - w_{\text{н}}), \tag{5}$$

где $w_{\text{н}}$ – новое значение веса; $w_{\text{м}}$ – старое значение; a – скорость обучения; x – величина входа.

Так как входной вектор x нормирован, т. е. расположен на гиперсфере единичного радиуса в пространстве весов, то при коррекции весов по этому правилу происходит поворот вектора весов в сторону входного сигнала. Постепенное уменьшение скорости поворота позволяет произвести статистическое усреднение входных векторов, на которые реагирует данный нейрон.

Как показало исследование, полученный алгоритм позволяет значительно ускорить работу программы голосовой идентификации. Данная модернизация позволяет использовать программу на предприятиях с большим потоком пользователей.

Также программный комплекс очень гибок и имеет большое пространство для дальнейшего усовершенствования и добавления новых функций, что делает его не только выгодным программным продуктом, но и перспективным проектом для развития и получения прибыли.

Литература

1. Bosi, M. Introduction to digital audio coding and standards / M. Bosi, R. E. Goldberg – Springer Science + Business, Media USA, 2003. – 434 p.
2. You, Y. AudioCoding: Theory and Applications / Y. You. – NY : Spring-er, 2010. – 349 p.
3. Загуменнов, А. П. Компьютерная обработка звука / А. П. Загуменнов. – М. : ДМК, 1999. – 384 с.